MOLECULAR AND CELLULAR BIOLOGY, Aug. 2007, p. 5921–5932 Vol. 27, No. 16
0270-7306/07/$08.00+0 doi:10.1128/MCB.00702-07

Sequence Dependence of Chromosomal R-Loops at the Immunoglobulin Heavy-Chain Sμ Class Switch Region▽†

Feng-Ting Huang,[1] Kefei Yu,[1] Barbara B. Balter,[3] Erik Selsing,[3] Zeliha Oruc,[4] Ahmed Amine Khamlichi,[4] Chih-Lin Hsieh,[2] and Michael R. Lieber[1]*

Departments of Pathology, of Biochemistry and Molecular Biology, of Molecular Microbiology and Immunology, and of Biological Sciences,[1] and Departments of Urology and of Biochemistry and Molecular Biology,[2] University of Southern California Keck School of Medicine, 1441 Eastlake Avenue, MC9176, Los Angeles, California 90089-9176; Department of Pathology, Tufts University School of Medicine, 150 Harrison Avenue, Boston, Massachusetts 02111[3]; and CNRS UMR 5089-IPBS, Equipe Instabilité Génétique et Régulation Transcriptionnelle, 205 route de Narbonne, 31077 Toulouse Cedex, France[4]

Received 20 April 2007/Returned for modification 29 May 2007/Accepted 1 June 2007

The mechanism by which the cytidine deaminase activation-induced deaminase (AID) acts at immunoglobulin heavy-chain class switch regions during mammalian class switch recombination (CSR) remains unclear. R-loops have been proposed as a basis for this targeting. Here, we show that the difference between various forms of the Sμ locus that can or cannot undergo CSR correlates well with the locations and detectability of R-loops. The Sμ R-loops can initiate hundreds of base pairs upstream of the core repeat switch regions, and the area where the R-loops initiate corresponds to the zone where the AID mutation frequency begins to rise, despite a constant density of WRC sites in this region. The frequency of R-loops is 1 in 25 alleles, regardless of the presence of the core Sμ repeats, again consistent with the initiation of most R-loops upstream of the core repeats. These findings explain the surprisingly high levels of residual CSR in B cells from mice lacking the core Sμ repeats but the marked reduction in CSR in mice with deletions of the region upstream of the core Sμ repeats. These studies also provide the first analysis of how R-loop formation in the eukaryotic chromosome depends on the DNA sequence.

Mammalian immunoglobulin (Ig) genes undergo two types of DNA recombination, in addition to a somatic hypermutation (SHM) process. V(D)J recombination occurs in early lymphocytes and assembles the variable-domain exon so that IgM can be made. Class switch recombination (CSR) occurs only at the Ig heavy-chain locus and is responsible for the change in the heavy-chain isotype from IgM to IgG, IgA, or IgE; this process is also called the heavy-chain isotype switch (6, 17, 29). CSR occurs at repetitive DNA elements called switch regions, which vary in sequence and length. All of the switch regions are more than 1 kb in length and consist of unit repeats of 25 to 80 bp. All are located downstream of a sterile transcript promoter, which is necessary for CSR. All have a G-rich nontemplate strand, and all are rich in sites at which a cytidine deaminase called activation-induced deaminase (AID) prefers to act, namely, WRC sites (37). The regional nature of CSR, which gave rise to the term regionally specific recombination (15), contrasts with the vast majority of other physiologic recombination systems in biology, which are regarded as site specific. The special features of switch regions (such as being long and repetitive and located downstream of a promoter, having G-rich nontemplate strands, and not having sequences conserved among the different switch regions or among vertebrates that carry out CSR) suggested that the mechanism would be unusual relative to other recombination processes in biology.

Investigators at the Honjo laboratory discovered the key lymphoid-specific enzyme for both CSR and SHM, AID (22, 23). AID is a 26-kDa protein which deaminates C in DNA (5, 25) but only when that DNA is single stranded (2, 26, 36, 38). A key question for CSR and SHM concerns how the DNA becomes single stranded. Because transcription appears to be required for both CSR and SHM, one may propose that the 9-bp bubble in the DNA (due to the RNA/DNA hybrid) created by the RNA polymerase is sufficient to serve this purpose. This simple explanation alone is not adequate because all transcribed genes in the genome would be mutated (as in SHM) or recombined (as in CSR). Therefore, there must be another explanation. Moreover, the switch regions evolved hundreds of million of years after the presence of AID because SHM is more ancient than CSR (1, 11, 32). Therefore, the unusual features of mammalian switch regions likely target these regions for recombination rather than mutation.

For CSR, we have previously proposed that R-loop formation can improve the availability of single-stranded regions in which AID can act (35, 37). Several groups, including ours, have demonstrated R-loop formation in switch regions in vitro when these regions are transcribed by prokaryotic polymerases (4, 13, 27, 31). In 2003, we demonstrated kilobase-length chromosomal R-loops in the murine Sγ3 and Sγ2b switch regions in vivo (35), and this finding has been confirmed (28). We note that R-loops are not the only mechanism by which AID can gain access to C's in duplex DNA. During SHM, transcription and single-stranded binding proteins (replication protein A in particular) may function to liberate lengths of single stranded-

* Corresponding author. Mailing address: Norris Comprehensive Cancer Center, Rm. 5428, University of Southern California, 1441 Eastlake Ave., MC9176, Los Angeles, CA 90089-9176. Phone: (323) 865-0568. Fax: (323) 865-3019. E-mail: lieber@usc.edu.

† Supplemental material for this article may be found at http://mcb.asm.org/.

▽ Published ahead of print on 11 June 2007.

ness (3). In addition, CSR in *Xenopus* species occurs in a region that is not G rich and does so at an efficiency that is only about fourfold lower than that in the mouse Sγ1 region (40). It is possible that the G-rich repeats of mammalian switch regions evolved to improve the generation of single-stranded regions by fourfold at each switch region. For a donor and an acceptor switch region, this improvement may be multiplicative (16-fold).

Previously, we were unable to carry out PCR across the Sμ region, and other groups have had similar difficulty (34). The Sμ, arguably the most important switch region, has remained unexamined up to this point. Mice with a deletion of the core Sμ region (designated ΔSμTR mice) retain surprisingly high levels of CSR (11 to 63%, depending on which acceptor switch region is utilized) (18). This finding posed a major challenge for any model of CSR targeting because the vast majority of the AGCT sites, the clusters of G's within the Sμ repeats, and the overall G richness were all eliminated. In contrast, the more extensive deletion in the Iμ exon-μ constant region exon (Cμ) deletion mice causes a reduction in CSR to a level about 50-fold lower that in the wild type, demonstrating that some major difference between the core Sμ and the Iμ-Cμ deletion mice must account for the efficiency of CSR (12), even though the core Sμ repeats are missing in both strains of mice. The difference between these two deletion strains (the difference between locations labeled 2 and 3 in Fig. S1A and between mouse alleles 2 and 3 in Fig. S1B in the supplemental material) must have some role in maintaining the efficiency of CSR. What could the difference between the remaining sequences be? Here, we found that the allele with the less severe deletion still permitted R-loop formation but that the larger deletion resulted in a marked reduction in R-loop formation. Therefore, a major fraction of R-loop formation can begin upstream of the core Sμ repeat region, and the large majoritiy of the R-loops initiate in a 50-bp region in which the nontemplate strand has a 50% G content. These observations help reconcile many aspects of CSR targeting at the Sμ and begin to establish DNA sequence rules for the initiation of chromosomal R-loops.

MATERIALS AND METHODS

Enzymes and reagents. All restriction enzymes were from New England Biolabs (Beverly, MA). Sodium bisulfite and other chemicals were from Sigma Chemical Co. (St. Louis, MO). The C57BL/6 mice were purchased from Jackson Laboratories (Bar Harbor, ME). The *Escherichia coli* RNase H1 gene was amplified from *E. coli* genomic DNA and cloned into the pTME plasmid (a kind gift from Carlos D. M. Filipe, McMaster University, Canada). The purification was done according to the previously published method (8), and the purified product was stored at a stock concentration of 2 mg/ml. The purified RNase H was carefully checked for even low levels of DNase or RNase activity and found to contain no exo- or endonucleolytic activity.

Purification and culturing of mouse splenic B cells. Mouse B cells were taken from spleens of 8- to 12-week-old C57BL/6 mice. Single-cell suspensions were prepared from spleens, and red blood cells were removed. Naïve B cells were purified by magnetic cell sorting (with a system from Miltenyi, Auburn, CA). A single-cell suspension was incubated with anti-CD43 magnetic beads for 15 min at room temperature. An LS magnetic separation column was placed onto the MACS MultiStand and washed with phosphate-buffered saline (PBS) containing 0.5% bovine serum albumin. The cell suspension and beads were loaded onto the LS column. The column was washed three times with 3 ml of PBS containing 0.5% bovine serum albumin, and the flowthrough was collected. The cell suspension was centrifuged for 10 min at 200 × *g* at 4°C. The naïve B cells were counted and cultured at a density of 2×10^5 cells/ml in the presence of 20 μg of lipopolysaccharide (LPS)/ml or LPS plus 5 ng of interleukin-4 (IL-4; R & D Systems Inc., Minneapolis, MN)/ml, as indicated below, for 2 days.

Extraction and purification of genomic DNA. Cell pellets were washed with PBS once and dissolved in 10 mM Tris–1 mM EDTA (pH 8.0). Sodium dodecyl sulfate (SDS) at a 0.5% final concentration and proteinase K at a 0.2-mg/ml final concentration were added. Cells were incubated at 37°C overnight. Phenol-chloroform extraction and ethanol precipitation of the genomic DNA were done. Genomic DNA was dissolved in 10 mM Tris–1 mM EDTA (pH 8.0) to 1 mg/ml. The genomic DNA was digested with EcoRV at 37°C overnight. Phenol-chloroform extraction and ethanol precipitation of DNA were repeated, and the DNA was dissolved in Tris-EDTA, pH 8.0.

RNase H was used to verify R-loop molecules in stimulated mouse B cells. As a control, the genomic DNA from stimulated mouse B cells was treated with *E. coli* RNase H before bisulfite treatment. First, the genomic DNA was digested with EcoRV for 5 h at 37°C. Second, 1 μg of RNase H was added to the same tube for 16 h at 37°C. Then another 1 μg of RNase H was added to the same tube for 6 h at 37°C. The treated genomic DNA was purified by phenol-chloroform extraction and ethanol precipitation. The DNA was dissolved in Tris-EDTA, pH 8.0, and was subsequently treated with bisulfite (see below).

Bisulfite modification assay. For the bisulfite treatment, 20 μg of digested genomic DNA in 30 μl was mixed with 12.5 μl of 20 mM hydroquinone and 457.5 μl of 2.5 M sodium bisulfite (pH 5.2). The mixture was covered with mineral oil in a microcentrifuge tube in the dark at 37°C overnight. The bisulfite-treated DNA was purified with the Wizard DNA cleanup system according to the protocol of the manufacturer (Promega, Madison, WI). The purified DNA was desulfonated with 0.3 M NaOH at 37°C for 15 min and recovered by ethanol precipitation. The purified DNA was resuspended in 30 μl of 10 mM Tris–1 mM EDTA (pH 8.0) and stored at −20°C.

The PCR was done for 30 cycles, the products were resolved on a low-melting-temperature agarose gel, and the correctly sized fragments were cut out from the gel and cloned using the TOPO-TA cloning kit (Invitrogen, CA). Plasmid DNA from each clone of interest was purified using the GenElute plasmid miniprep kit (Sigma, St. Louis, MO). Sequencing reactions were carried out using the SequiTherm Excel II sequencing kit (Epigenetics, Madison, WI) and a model Primus 96 Plus thermal cycler (MWG Biotech, High Point, NC). Automated sequencing was carried out using the model 4200 DNA analyzer (Li-Cor, Lincoln, NE).

Controls to ensure that the results of the bisulfite modifcation assay reflected an in vivo structure included the following. Pretreatment of the genomic DNA with RNase A prior to the bisulfite treatment did not result in a reduction in the R-loop detectability. Mock incubations (without the RNase H) did not affect R-loop detectability. R-loops were never detected in untranscribed zones. Unstimulated B cells do not have any detectable R-loops in downstream switch regions (35).

DNA methylation assay. For the DNA methylation assay, 10 μg of digested genomic DNA was denatured with 0.3 M NaOH at 37°C for 15 min. The denatured genomic DNA was mixed with 12.5 μl of prewarmed (55°C) 20 mM hydroquinone and 457.5 μl of 2.5 M sodium bisulfite (pH 5.2). The mixture was covered with mineral oil in a microcentrifuge tube in the dark at 55°C for 4 h. The bisulfite-treated DNA was purified with the Wizard DNA cleanup system according to the protocol of the manufacturer (Promega, Madison, WI). The purified DNA was desulfonated with 0.3 M NaOH at 37°C for 15 min and recovered by ethanol precipitation. The purified DNA was resuspended in 20 μl of 10 mM Tris–1 mM EDTA (pH 8.0) and stored at −20°C.

Oligonucleotides. Oligomers were from Operon (Richmond, CA). The following primers were used in the PCR. For the amplification of the core Sμ from the C57BL/6 mouse strain, FTH84, a "converted" primer with a sequence in which three C's had been converted into three T's (5′-TGAGTTGGGGTAAGTTGGGATGAGT-3′); KY293, a native primer (5′-AACTCTACTGCCTACACTGGAC-3′); and/or KY294, a native primer (5′-CAGCACAATCTGGCTCACTT-3′), were used in the combinations specified below. For the amplification of the upstream and downstream regions of the core Sμ, FTH111, a converted primer with a sequence in which three C's had been converted into three T's (5′-AGATAAGTTAGGTTGAGTAGGGTT-3′); FTH119, a converted primer with a sequence in which three C's had been converted into three T's (5′-CTATTCTTTCTCAATTCTATACAACTA-3′); FTH50, a native primer (5′-TTGAAGGAACAATTCCACACAAA-3′); and FTH94, a native primer (5′-CTGGGAGAACTATTCTCATCCCAAA-3′), were used in the combinations specified below. For the amplification of the Iμ-Cμ from the mouse strain with the larger Sμ deletion, FTH52, a converted primer with a sequence in which six C's had been converted into six T's (5′-AATGGTAAGTTAGAGGTAGTTAT-3′); FTH51, a native primer (5′-CCCATGGCCACCAGATTCTTATC3′); and FTH48, a native primer (5′-TCTCCATTCAATTCTTTTCCAATA-3′), were used in the

combinations specified below. For the probe in the hybridization in the core Sμ region from the C57BL/6 mouse strain, FTH98 (5′-ACTCAACTCAACTCAA CTCAACTCAA-3′), FTH102 (5′-AATTCTAACCAACCAACTCTACTCA-3′), FTH104 (5′-ACTCAACTCAACTCAACCCAACTCAA-3′), and FTH105 (5′-ACCCAACTCAACCCAACTCAACCCAA-3′) were used as a radiolabeled mixture. For the probe in the hybridization for the ΔSμTR mouse strain, FTH130 (5′-ATACAACTATAACCTTCCTTCTACAT-3′), FTH131 (5′-CACA TTAAATTATAAATCAAAAATATAATAA-3′), FTH132 (5′-CATCAACCA ACCCAATTAAATCCAA-3′), FTH134 (5′-ACTCAACCCAATTCATAATCC CAAT-3′), and FTH135 (5′-ATATAAATAACCCAAACAACAATACTC-3′) were used as a radiolabeled mixture. For the probe in the hybridization for the Iμ-Cμ deletion mouse strain, FTH178 (5′-ACATATAAACTAACTTAAAAAC CCTTC-3′), FTH179 (5′-AAAAAACCCAAAATCCAAACCTACC-3′), and FTH180 (5′-CTTTAAAAACAACAACCACAACTATAA-3′) were used as a radiolabeled mixture.

Determination of the frequencies of R-loops. In the first phase of the method to determine the frequencies of R-loops, the bisulfite-treated, stimulated B-cell genomic DNA was amplified using two native primers. The PCR was carried out for 30 cycles, the products were resolved on a low-melting-temperature agarose gel, and the correctly sized fragments were cut out from the gel and cloned using the TOPO-TA cloning kit (Invitrogen, CA). Each white clone was picked and restreaked as a line onto the surface of a new ampicillin agar plate. Each plate contained 110 different clones. The PCR products included molecules in which the top (nontemplate) strand of the original bisulfite-treated molecule served as the template or in which the bottom (template) strand did so. By reading the sequences at sites of sporadic bisulfite conversion in 32 randomly chosen clones, the ratio of top- and bottom-strand molecules was determined in order to permit estimates of what fraction of the total population analyzed was informative about the top strand.

The second phase of the method involved colony lifts onto nylon membranes. The nylon membrane was pressed against the agar plate (containing 110 clones) for 2 min. The membrane was then transferred into a denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 15 min with the bacterial side up. The membrane was then transferred into 1 M Tris, pH 7.5, for 15 min, followed by 0.5 M Tris (pH 7.5)–1.5 M NaCl for 15 min, followed by a rinse with 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate). Kimwipes were used to wipe the bacterial debris off the surface of the membrane while the membrane was under the 2× SSC. The DNA was then fixed onto the membrane by UV cross-linking. The membrane was then rinsed with 2× SSC and put into a plastic hybridization bag.

In the third phase, the converted primers were individually 5′-end labeled with polynucleotide kinase and then mixed to generate the probes for hybridization. Each membrane was prehybridized at 66°C for 15 min. The oligonucleotide probes (5 pmol each) were 5′-end labeled using polynucleotide kinase and then added, and the membrane was hybridized at 66°C overnight. The membrane was washed with 2× SSC–0.5% SDS for 20 min twice and 0.1× SSC–0.5% SDS for 20 min twice. The membrane was exposed overnight. The clones corresponding to positive signals were confirmed by DNA sequencing.

The same membranes were hybridized with probes specific to converted template-strand molecules. No template-strand molecule having long stretches of conversion was detected.

RESULTS

R-loops at murine core Sμ. The bisulfite modification assay can be used to obtain structural information about any non-B DNA structure that has regions of transient or stable single strandedness (35, 39). The bisulfite anion carries out a nucleophilic attack on the bond between C-5 and C-6 of cytosine only when the bond is unprotected by stacking, which occurs when the DNA is single stranded. The reactivity of all of the C's in a region indicates a stable zone of single strandedness. An enrichment method using one converted or enriching primer (in which the C-to-U changes are anticipated) and one conventional (native) primer for the PCR can be used to enrich with R-loops.

Though we previously documented R-loops at the Sγ3 and Sγ2b regions, the Sμ region is the donor switch region for recombination with all of the acceptor switch regions (the Sγ, Sα, and Sε regions) and, hence, is the most important. We tested for R-loops at the murine Sμ by using an enriching primer-native primer PCR pair after bisulfite modification (Fig. 1A; also see Table S1 in the supplemental material). The expected PCR product size was 1,253 bp. Because of the repetitive nature of the core Sμ, the converted primer can prime at multiple locations. Therefore, the products of the PCR included multiple species, most of which were shorter than the full-length Sμ; hence, a single product band was not obvious at the full-length position. The region surrounding the position of the anticipated full-length DNA was cut out, cloned, and used to transform *E. coli*.

Upon sequencing, molecules having different lengths of long stretches of single strandedness were observed (Fig. 1A). The longest was nearly 1,100 nucleotides. The majority of R-loops ended within the core Sμ. Six of 34 molecules extended downstream of the core Sμ, and their end points fell within the region containing several degenerate Sμ repeats. The sequenced molecules had deletions of different sizes. In some cases, the deletion product was several hundred base pairs shorter than the expected product. The reason for the deletions was that the core Sμ region consists of uniform direct repeats and the converted primer can anneal at many regions within the core Sμ, not simply at the designated locations. In addition, the Sμ sequence may be unstable in bacteria. Other groups have previously reported deletions within the Sμ sequence (24).

RNase H treatment was used to assess whether the long stretches of single strandedness were due to R-loop formation. The RNase H destroys the RNA in R-loops, thereby permitting Sμ repeats in the top (nontemplate) strand to anneal to repeats in the bottom (template) strand. If this annealing occurs, then we expect to lose the long stretches of bisulfite reactivity and the enriched primer can no longer anneal to a site that is not converted. That is, the destroyed R-loops are not amplified by the enriching primer after the RNase H treatment. One gapped molecule was detected, which had discontinuous stretches of single strandedness. It consisted of several alternations between duplex DNA and single-stranded DNA. This finding most likely reflects evidence of incomplete action by RNase H, but it may also reflect a misalignment of the repeats (35) (see Discussion).

Frequency of R-loops at murine core Sμ. To determine the frequency of R-loops at the Sμ allele in DNA from splenic B cells that were stimulated with LPS–IL-4, the core Sμ was first amplified from bisulfite-treated genomic DNA by using a pair of native primers, which avoids any enrichment. The native primers were close to the core Sμ. The expected PCR product size was 1,518 bp, but the actual PCR product sizes were distributed over a range around 1,518 bp, and this region was used for cloning and the transformation of *E. coli*. Colony lift hybridization was done using a converted oligonucleotide probe which anneals to any bisulfite-reactive form of the top strand in any R-loop. The probe was actually a mixture of four oligonucleotide probes, corresponding to four different locations within the Sμ core region, so that R-loops covering a larger fraction of the switch region could be detected.

After testing 921 colonies (see Materials and Methods), 296 were determined to provide information about the top (nontemplate) strand, whereas the remainder reflected the bottom

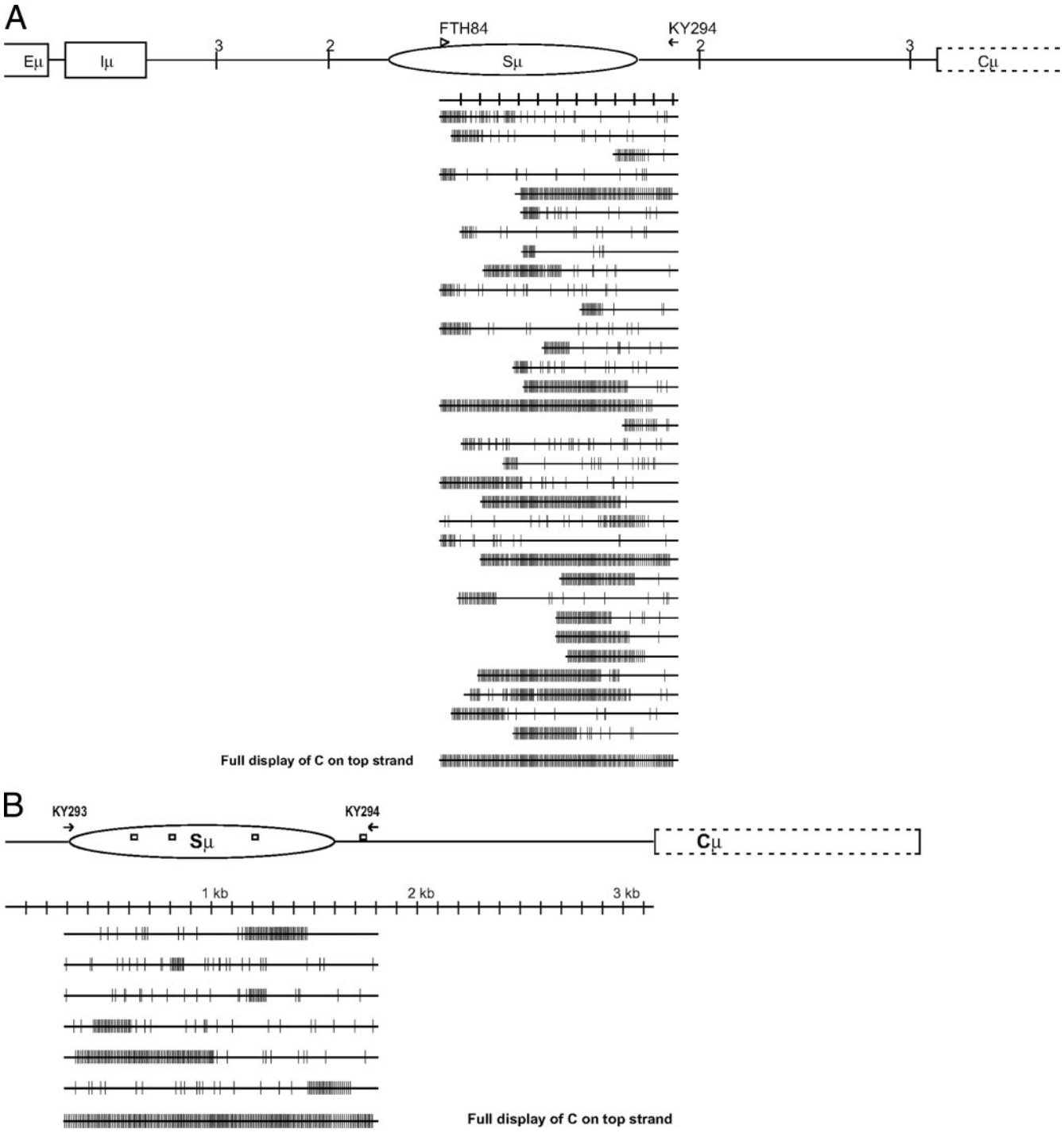

FIG. 1. R-loop molecules at the murine core Sμ in stimulated B cells from C57BL/6 mice. (A) Single strandedness on the top strand in the murine core Sμ in stimulated B cells from C57BL/6 mice. After bisulfite treatment, the region was amplified using one native and one converted primer. The converted primer (shown as an open arrowhead), FTH84 (which starts at nucleotide 137610 of the sequence corresponding to GenBank accession no. AC073553), with a sequence in which three C's were converted into three T's, was located 250 bp downstream of the start of the core Sμ. The native primer, KY294, was located 200 bp downstream of the end of the core Sμ. The expected PCR product was 1,253 bp, but the actual PCR products ranged from this size to smaller sizes. The region of the gel corresponding to approximately 1,253 bp was cut out and cloned. Thirty-three clones showed different lengths of R-loops in the core Sμ. All clones had small deletions and mutations. Among all the clones, 25 clones had large deletions, and these clones are shown as shorter lines. All large deletions occurred in the core Sμ region. The naive B cells were stimulated with 20 μg of LPS/ml for 2 days. In the diagram at the top, the long ellipse represents the core Sμ region. The vertical bars labeled 2 indicate the boundaries of the core Sμ deletion, and the vertical bars labeled 3 indicate the boundaries of the larger deletion in the Sμ allele in the knockout mouse strains. Eμ, intronic enhancer. The dashed lines in the Cμ segment indicate exon/intron boundaries that are not specified. The line at the bottom of the panel displays every C on the top strand of the PCR product. (B) R-loop molecules at the murine core Sμ in stimulated B cells from C57BL/6 mice. Primary splenic B cells were stimulated in culture for 2 days with LPS–IL-4. The genomic DNA was

(template) strand. Colony lift hybridization showed that 12 of the corresponding molecules had long stretches of conversion consistent with R-loops (Fig. 1B; duplicate molecules are not shown). Hence, the R-loop frequency was approximately 4% of alleles (12 out of 296). Of course, this finding may be an underestimate of the actual R-loop frequency because the probes covered only a small fraction of the region. Therefore, R-loops between or upstream or downstream of the probe-covered regions may exist. More importantly, R-loops that extend to the site where the native primer is located will not be amplified because the native primers will not anneal at priming sites that have had their single-stranded C's converted by bisulfite. In other words, efficient native-primer amplification rests on the assumption that the native-primer site is in a duplex DNA conformation. For these and other reasons, this estimate of frequency may be an underestimate.

Boundaries of the R-loops at murine Sμ in wild-type mice and in mice with the core Sμ deletion (ΔSμTR). CSR breakpoints are not restricted to the core Sμ, in contrast to most breakpoints in the acceptor switch regions (6). In the Sμ region, somewhat more than half of the breakpoints are in the core Sμ (see Fig. S1 in the supplemental material). About one-third are upstream of the core Sμ, and about 8% are downstream of the core Sμ. This raises the possibility that the R-loops in the Sμ region may not be restricted to the core repeats of the Sμ. To determine the boundaries of the R-loops in the Sμ, converted primers for the upstream or downstream regions of the core Sμ were designed. For the upstream region of the core Sμ, the PCR product was 770 bp. Upon cloning and sequencing, molecules having different lengths of long stretches of single strandedness were detected (Fig. 2B). The shortest one was 54 nucleotides long, and the longest one was 581 nucleotides. For the downstream region of the core Sμ, the PCR product was 721 bp. Four molecules having long stretches of conversion were detected (Fig. 2D).

RNase H treatment was done to confirm that these molecules were R-loops, and the treatment was done in the same way as that for the core Sμ. After the bisulfite modification assay and PCR, no PCR product corresponding to the upstream region of the core Sμ was detectable, indicating that the molecules were, indeed, R-loops and, therefore, could not be amplified by the converted primer after RNase H treatment (Table 1). For the region downstream of the core Sμ, the PCR product was cut out and cloned. After the sequencing of 23 molecules, no molecule was found to have long stretches of conversion. The fact that the long, single-stranded DNA regions were eliminated by RNase H documents that these regions were part of an R-loop.

The same primer sets were used for the core Sμ deletion (ΔSμTR) mouse strain to detect any R-loops in the stimulated B cells. At the core Sμ deletion alleles, the CSR efficiency was only about one-third that in the wild-type mice (18) (see Fig. S1 in the supplemental material). (The efficiency varies depending on which acceptor switch region is involved.) R-loops upstream and downstream of the former location of the core Sμ region were detected (Fig. 2A and C), and these R-loops were indistinguishable from the R-loops found in the corresponding regions of the wild-type allele mentioned above. The R-loop nature of these molecules was confirmed by RNase H treatment (Table 1). Therefore, R-loops can initiate upstream and can extend downstream of the core Sμ region.

Frequency of the R-loops at murine Sμ in the core Sμ deletion (ΔSμTR) B cells. To determine the frequency of R-loops in the ΔSuTR murine B cells, a pair of native primers was designed to amplify the Iμ-Cμ from the stimulated B cells of the ΔSuTR mice. One primer corresponded to the Iμ exon, and the other one corresponded to the Cμ. The PCR product was 2,251 bp. The composite probe for hybridization, which was specific for converted top-strand molecules, contained five different oligonucleotide probes: three were positioned at different locations in the region upstream of the former core Sμ, and two were positioned downstream. After testing 932 colonies (see Materials and Methods), 311 were determined to provide information about the top (nontemplate) strand, whereas the remainder reflected the bottom (nontemplate) strand. Among these colonies, 11 molecules having long stretches of conversion were identified (Fig. 3). Therefore, the frequency of R-loops in the ΔSμTR allele at day 2 of IL-4–LPS stimulation was 3.5% (11 out of 311).

RNase H treatment was done to test whether molecules detected by the filter hydridization were indeed R-loops. We found that RNase H treatment destroyed 92% of the molecules that were otherwise detected by the single strand-specific probes (see Materials and Methods). Therefore, nearly all of these long stretches of conversion were attributable to R-loops.

The filter membranes were also hybridized with oligonucleotide probes specific for converted bottom (template)-strand molecules, but no molecules having long stretches of conversion were detected.

When we analyzed the data from the ΔSμTR locus (Fig. 3), we found a consistent, but slightly variable, lack of conversion in the DNA at the center of the R-loops within the boundaries of the area in which the knockout deletion occurred (deletion boundaries designated by the number 2 in Fig. 3) in 10 of 11 molecules. We were initially quite perplexed by this finding because we felt that the R-loops should extend right through this 93-bp region of prokaryotic DNA. Some of the sites that failed to react with bisulfite were CpG sites, and DNA methylation of the C in these sites could make them unreactive with bisulifte. Though this could explain some of the instances in

prepared and then treated with sodium bisulfite as described in Materials and Methods. A single round of PCR (30 cycles) was done using two regular (native-sequence) primers, KY293 and KY294 (shown as black arrows). The PCR product was 1,518 bp and was cloned into *E. coli*. About 1,106 clones were transferred onto a filter membrane by colony lifting. After hybridization with a probe, 12 positive clones were identified. The sequences of these clones are represented here, and these sequences show long stretches of single strandedness, the longest one being 650 nucleotides. In the diagram at the top, the long ellipse represents the core Sμ region, which is 1,287 bp long. The region between the core Sμ and the Cμ is 1,535 bp. The open square in the diagram indicates the location of the probe. Each long line represents an independent molecular clone. The small vertical bar on each line indicates a position in the sequence at which a C has been converted into a T. The bottom line displays every C on the top strand of the PCR product.

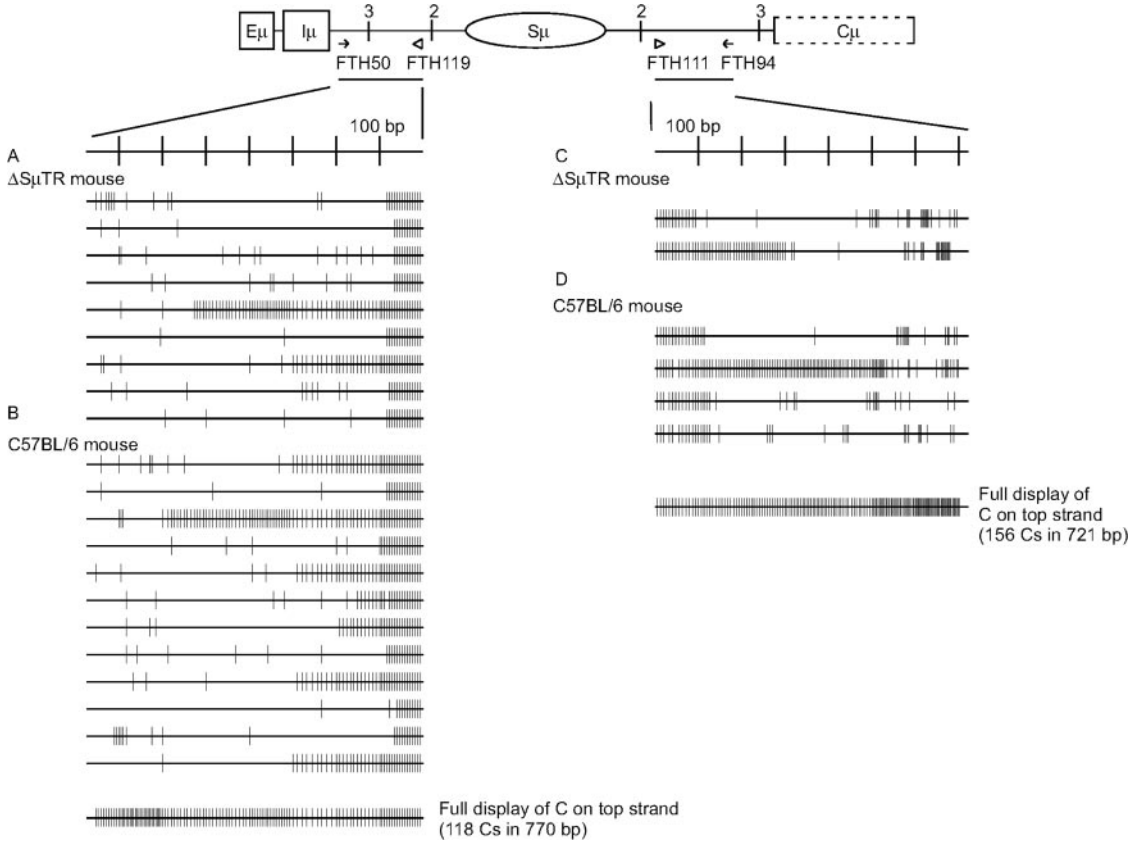

FIG. 2. Single strandedness on the nontemplate strand upstream and downstream of the murine core Sμ in stimulated B cells. (A) Top (nontemplate)-strand sequence at the upstream region of the core Sμ in the core Sμ deletion (ΔSμTR) mouse strain. The genomic DNA was derived from B cells (from the ΔSμTR mouse strain) stimulated for 2 days with LPS–IL-4. A single round of PCR (30 cycles) was done using one regular (native-sequence) primer and one converted primer (shown as an open arrowhead) whose sequence was complementary to the top strand in which the C's were converted into U's. FTH119, the converted primer, with a sequence in which three C's were converted into three T's, was located 400 bp upstream of the core Sμ. FTH50, the native primer, was located 50 bp downstream of the Iμ exon. The diagram at the top presents a map of the Ig μ gene of the mouse strain. Eμ, intronic enhancer. The vertical bars labeled 2 indicate the boundaries of the core Sμ deletion (ΔSμTR). The vertical bars labeled 3 indicate the boundaries of the larger deletion in the Sμ allele. All symbols are the same as those in Fig. 1. (B) Top-strand sequence at the upstream region of the core Sμ in the C57BL/6 mouse strain. The genomic DNA was the same as that used for Fig. 1. (C) Top-strand sequence in the downstream region of the core Sμ in the core Sμ deletion (ΔSμTR) mouse strain. A single round of PCR (30 cycles) was done using one regular (native-sequence) primer (arrows) and one converted primer (open arrowheads). FTH111, the converted primer, with a sequence in which three C's were converted into three T's, was located 440 bp downstream of the core Sμ. FTH94, the native primer, was located 400 bp upstream of the Cμ. (D) Top-strand sequence of the downstream region of the core Sμ in the C57BL/6 mouse strain.

which these C's were unreactive, it could not explain all aspects of the observed pattern of reactivity and unreactivity. We then realized that the palindromic *loxP* site is located in the middle of this region. The palindromic nature of the *loxP* site permits it, when single stranded, to form a 13-bp stem and an 8-nucleotide terminal loop (see Fig. S2 in the supplemental material). The C's in the loop always reacted with bisulfite, but three C's in the stem did so only rarely. Two of these three C's in the stem were complicated by possible CpG methylation, but the remaining one was not a CpG site, and the low level of reactivity is best explained by base stacking in the stem of the stem-loop. A DNA methylation assay confirmed that the nonreactivity of that C in the stem of the stem-loop could not be accounted for by DNA methylation (see Fig. S3 in the supple-

TABLE 1. Single strandedness was destroyed after in vitro RNase H treatment[a]

Region of single strandedness	No. of molecules with long stretches of conversion/total no. of sequenced molecules for:			
	C57BL/6 mouse samples without RNase H treatment	C57BL/6 mouse samples with RNase H treatment	ΔSμTR mouse samples without RNase H treatment	ΔSμTR mouse samples with RNase H treatment
Upstream of core Sμ	13/18	ND (no PCR product)	9/31	ND (no PCR product)
Core Sμ	34/119	0/2	NA	NA
Downstream of core Sμ	4/31	0/23	2/32	0/32

[a] The RNase H treatment removes the RNA in the R-loops. The cutoff length for long stretches of conversion was 50 nucleotides. ND, not detectable; NA, not applicable.

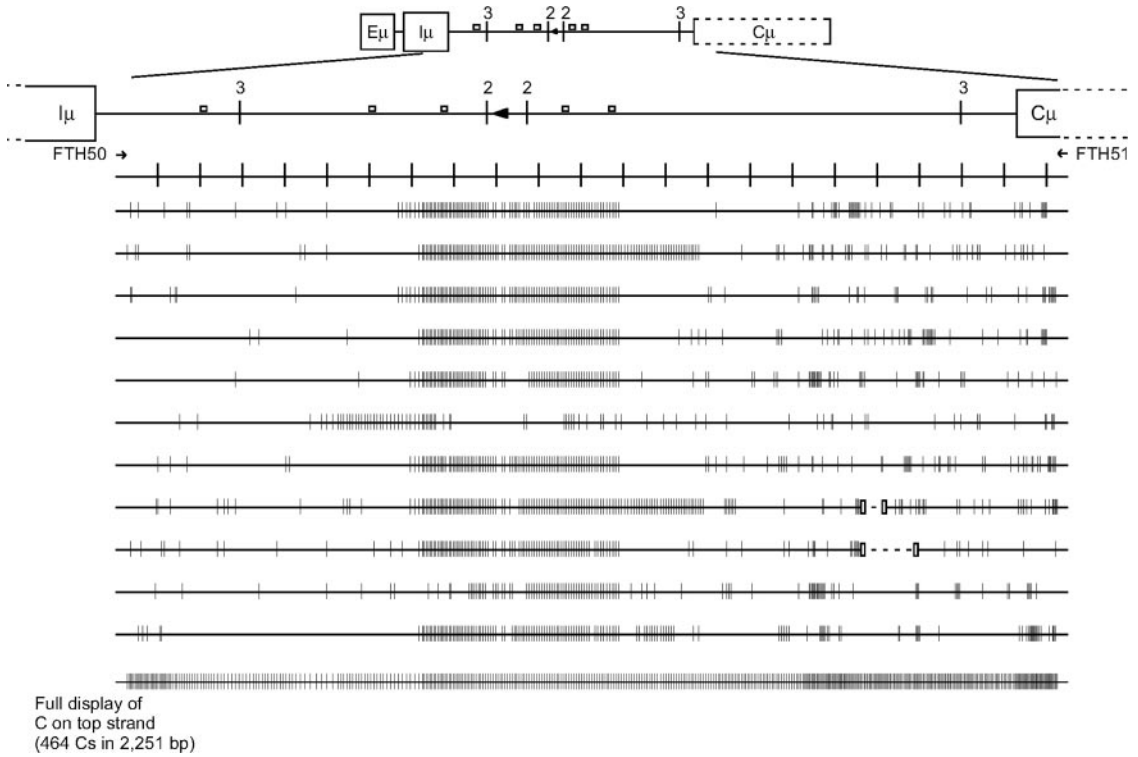

FIG. 3. R-loop molecules at the murine Iμ-Cμ in stimulated B cells from the core Sμ deletion (ΔSμTR) mouse strain. The genomic DNA was the same as that used for Fig. 2. After the bisulfite treatment, a single round of PCR (30 cycles) was done using two (native-sequence) primers, FTH50 and FTH51. The PCR product was 2,251 bp and was cloned into *E. coli*. About 932 clones were put onto a filter membrane by colony lifting. After probe hybridization, 11 positive-signal clones having long stretches of conversion were identified. The diagram at the top shows the map of the Iμ-Cμ of the ΔSμTR allele. The arrowhead in the diagram indicates the *loxP* sequence at this allele. The dashed lines indicate that there is no sequence information for those portions of the clones. The vertical hollow boxes mark the separation of the dashed lines from the solid lines. All symbols are the same as those in Fig. 1. Eμ, intronic enhancer.

mental material). Moreover, two other C's were substantially more unreactive than their methylation status could account for. This lack of reactivity in the three C's of the stem is quite consistent with the protection of the stem region from bisulfite reactivity in the native bisulfite modification assay for assessing single strandedness. These results provide an independent indication that this region in the middle of the nontemplate DNA strand of the R-loop is single stranded. This strand can form secondary structures that interfere with bisulfite reactivity by permitting stacking of the C's, but this formation is not usually observed, except, for example, in palindromic regions like the one described here. The lack of such interruptions in a majority of R-loops that extend through this region on the wild-type allele is most consistent with a lack of any other secondary structure (such as a G quartet).

R-loops in alleles from mice with larger deletions of the core Sμ and the surrounding sequences were not detected. A larger region between the Iμ exon and the Cμ was previously removed for a different knockout allele in mice (12). In the B cells of these mice, the CSR efficiency was only about 2% of that seen in B cells from the wild-type mice (12) (see Fig. S1 in the supplemental material). (The efficiency varies depending on the acceptor switch region.) The converted primer FTH52, with a sequence in which six C's have been converted into T's, was designed to test for the presence of R-loops in the B cells of the deletion mice. FTH52 is a strongly enriching primer because of the large number of C's converted into T's. The PCR product is 505 bp. After the sequencing of 14 molecules, no molecules containing long stretches of conversion were found, despite the use of the very strongly enriching primer (Fig. 4). Therefore, no R-loops at alleles that had a more extensive deletion surrounding the Sμ were detected.

Although no R-loops were detected by the enrichment method, the colony lift hybridization was done to further search for any R-loops at the allele with larger deletions. To avoid missing any R-loops at the upstream region, where the loci of CSR breakpoints in B cells of mice with the deletion allele were located, the PCR product was amplified using one more upstream native primer, which corresponded to the Iμ exon, paired with the same downstream native primer used for the ΔSuTR mice. The PCR product was 1,018 bp. The composite probe for hybridization, which was specific for converted top (nontemplate)-strand molecules, contained three different oligonucleotide probes. After testing 1,029 colonies (see Materials and Methods), 450 were determined to provide information about the top strand, whereas the remainder reflected the bottom (template) strand. Among these colonies, no molecules containing long stretches of conversion were detected. Therefore, the frequency of R-loops at this allele was below 1 per 450.

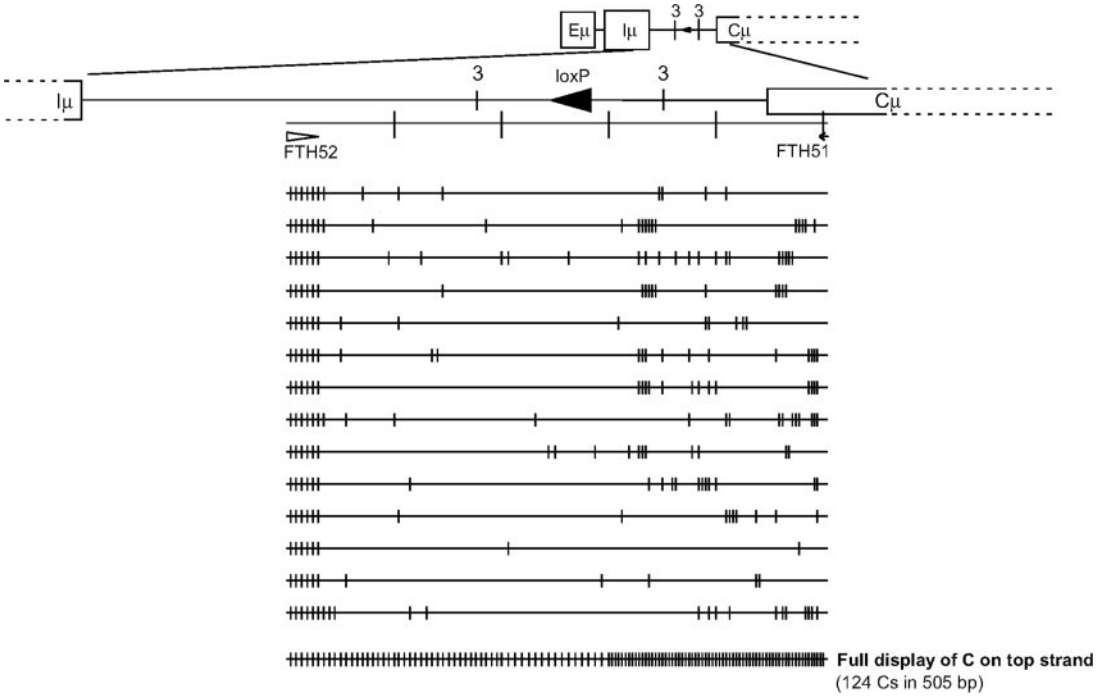

FIG. 4. Test for single strandedness on the nontemplate strand in the murine Iμ-Cμ in the stimulated B cells from mice having the larger deletion around the core Sμ. The genomic DNA was obtained from B cells, stimulated for 2 days with LPS–IL-4, from the mice with the larger deletion around the core Sμ. A single round of PCR (30 cycles) was done using one regular (native-sequence) primer and one converted primer. FTH52, the converted primer, with a sequence in which six C's were converted into six T's, was located 200 bp downstream of the Iμ exon. FTH51, the native primer, was 70 bp downstream of the start of the Cμ. The PCR product was 505 bp, including the 175-bp heterologous DNA from the construct for making the deletion allele. Among 14 different clones, no molecule had long stretches of conversion. All symbols are the same as those in Fig. 1. Eμ, intronic enhancer.

DISCUSSION

In vivo R-loop formation at prokaryotic replication origins (19), mitotic recombination hot spots in *Saccharomyces cerevisiae* (10), avian G-rich sequences (under specific circumstances) (14), and Ig class switch regions (9, 30, 35, 41) has been described previously. Despite these key findings in various systems, the in vivo sequence determinants of R-loop formation have yet to be explored. The studies here provide an initial set of functionally relevant sequences that help delimit what regions do and do not form R-loops in vivo.

Sites of R-loop initiation, of CSR, and of AID deamination. R-loops at the ΔSμTR allele were found at a frequency similar to that at the Sμ in the wild-type mouse strain. No R-loops in the Iμ-Cμ deletion mouse strain were detected. Therefore, the sequences upstream of the core Sμ repeats can be important sites for initiating R-loops. These findings correlate well with the findings on CSR efficiency (18) (12), which imply that the R-loops are the targets in the CSR. For the wild-type and the ΔSμTR loci, respectively, in B cells stimulated with LPS–IL-4, we found a frequency of 1 R-loop per 25 and per 28 alleles by using native primers (no enrichment). The agreement of these numbers is encouraging, given that the native primers used were different. (The primers for the wild-type allele must be adjacent to the core Sμ repeats to detect a PCR product.)

The results of this study indicate a substantial correlation between the locations of R-loops, the locations of recombination breakpoints, and the locations of AID sites of C-to-U conversion (Fig. 5 and 6). Though we believe that many R-loops initiate within the core Sμ repeats, unexpectedly, we find that many R-loops can begin upstream of the core Sμ repeats (Fig. 2). In fact, we think that the majority (but not all) of R-loops initiate in the region upstream of the core Sμ, because there was not much difference in the R-loop frequencies between wild-type and ΔSμTR alleles.

What may the mechanism for the initiation of R-loops at these upstream locations be? R-loops that begin upstream of the Sμ do so at clusters of G's. We do not know the precise combination of clustered G's that are important for the initiation of R-loop formation, but we know that a relatively random DNA sequence does not cause R-loop formation in the mammalian genome. For example, even the most strongly enriching converted primers (with seven C-to-T conversions over 22 nucleotides) did not permit the detection of R-loops in the Cγ3 region under conditions in which R-loops in the Sγ3 region could be readily detected. Moreover, even when two enriching (converted) primers at the Cγ3 region were used to achieve even greater sensitivity, no single strandedness on either the template or the nontemplate strand was detected. Hence, R-loop formation does not occur in the constant regions, despite their being in the same transcription unit. Therefore, specific patterns and lengths of G clustering increase the probability of R-loop formation, and deviations from those patterns make R-loops increasingly less likely.

The majority of R-loops that initiate upstream of the Sμ

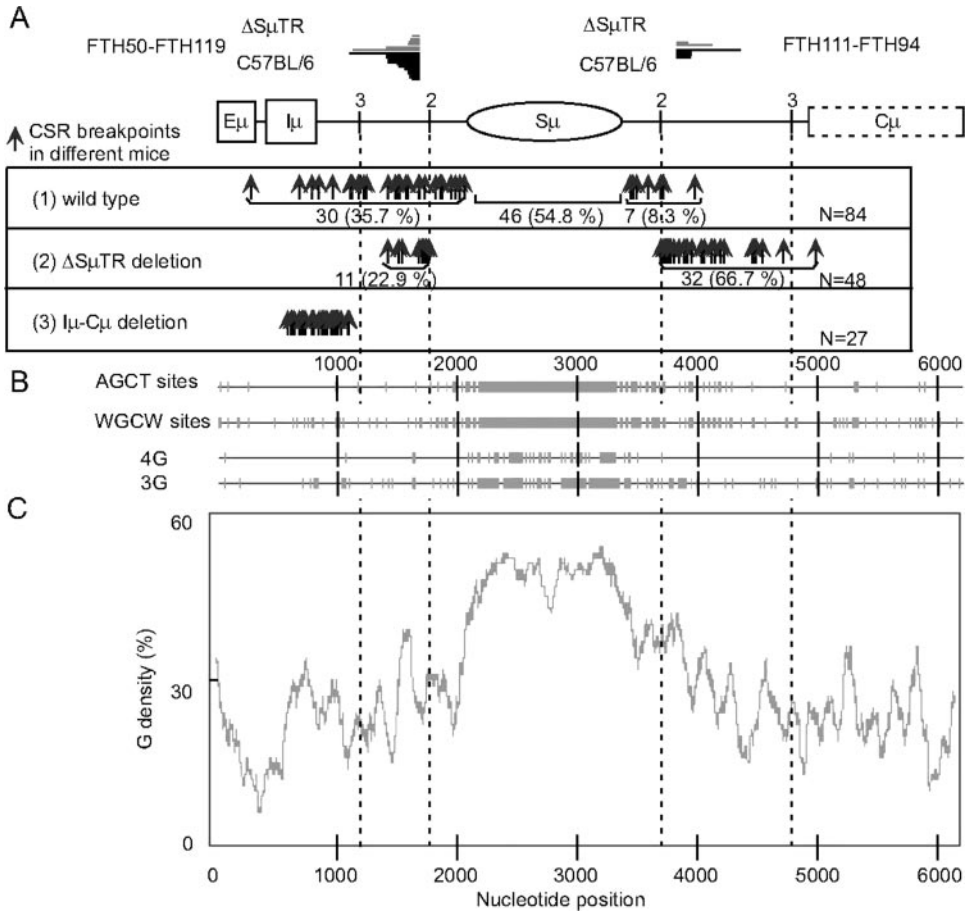

FIG. 5. Plots of R-loop locations, CSR breakpoints, AGCT sites, and G densities for the murine Sμ through the initial portion of the first constant exon. (A) The relative positions of the 313-bp intronic enhancer (Eμ), the 454-bp Iμ exon, the 1,205-bp Iμ-Sμ intervening region, the 1,287-bp Sμ, the 1,535-bp Sμ-Cμ intervening region, and the first 1,287 bp of the Cμ (exon/intron boundaries are not specified) are shown. R-loop upstream and downstream boundaries based on the data in Fig. 2 are shown as lines at the top of the diagram (dark lines are wild type, and grey lines are ΔSμTR). The locations of CSR breakpoints in different mice are shown as arrows. (B) Plot of the AGCT, WGCW, four-G cluster, and three-G cluster sites of the sequence in panel A. (C) Plot of the G densities of the sequence in panel A.

begin at a specific site with the sequence GGGGCTGGGG, which is within a 50-bp zone that has a 50% G content on the nontemplate strand (Fig. 2, 3, and 6). Interestingly, this region is also close to the area where the top strand begins to show a distinct increase in AID-mediated C-to-U deaminations (Fig. 6). In a recent paper from the Neuberger laboratory, it was unclear what could account for the increase in AID mutation frequency at this region because the densities of AGCT sites (and other preferred WRC sites) did not increase until several hundred base pairs downstream (34). The peak of G density and the clusters of GGGG or GGG in this region where most of the upstream R-loops initiate appear to account for the sharp rise in AID sites of action here (Fig. 5 and 6) (34).

Among R-loops detected in wild-type B cells or B cells with the core Sμ deletion (ΔSμTR), nine R-loops began even further upstream than those in the region demarcated in Fig. 6. Seven of these initiated within a region with clusters of three G's, and the remaining two also initiated near several clusters of three or four G's. The latter two initiated upstream of the Iμ-Cμ deletion boundary. Rare R-loops initiating here may explain the low but distinct occurrence of CSR events that have breakpoints within the residual region between the Iμ exon and the Cμ in the Iμ-Cμ deletion B cells (12).

The core repeats are not essential for R-loop formation (this study) or for CSR (18). The initiation of R-loops upstream of the core repeats may be evolutionarily advantageous because it permits a greater fraction of the zone of WRC sites in the core Sμ repeats to be included in a zone of single strandedness. Of course, the GGGGT portions of the Sμ repeat can also initiate R-loops, accounting for the R-loops initiating within the core Sμ.

In the Sμ core deletion mice (Fig. 3), some level of residual CSR occurs upstream and downstream of the position of the deletion boundary (20, 21), and the ones occurring upstream in the core Sμ deletion mice would be particularly difficult to explain if R-loops initiated only in the core Sμ region. Because we now know that the R-loops can begin upstream, the breakpoints upstream of the core Sμ in the deletion mice can now be more easily understood. The CSR breakpoints downstream of the core Sμ are not difficult to explain, given that R-loops extend past the repetitive zone (Fig. 2C and D).

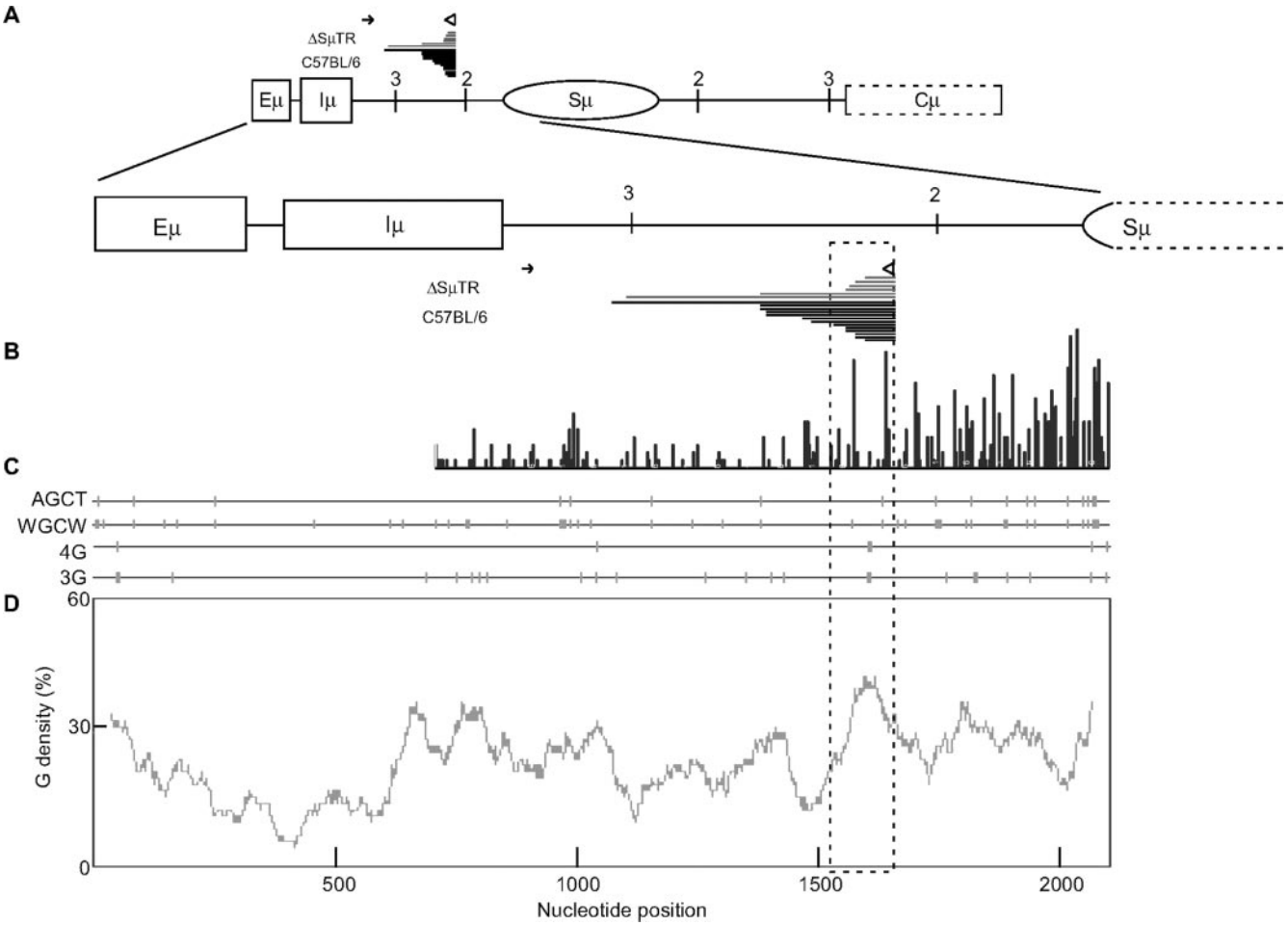

FIG. 6. Plots of R-loop locations, mutation frequencies, AGCT sites, and G densities for the region upstream of the core Sμ. (A) Magnification of the region upstream of the core Sμ depicted in Fig. 5. Each line represents different R-loops initiating at different places upstream of the core Sμ. Eμ, intronic enhancer. (B) The mutational frequencies for the corresponding region from the *msh2*$^{-/-}$ *ung*$^{-/-}$ mice as published by others (34). (Note that only the top [nontemplate]-strand mutations [provided by M. Neuberger] are shown, rather than the composite top [nontemplate]- and bottom [template]-strand mutation data that were published previously [34].) Without the detailed data from Xue et al. (34), the R-loop data in the present paper permitted a prediction of where the nontemplate-strand mutation frequency would start to increase. (C) Plot of the AGCT, WGCW, four-G cluster, and three-G cluster sites of the sequence in panel A. (D) Plot of the G densities of the sequence in panel A. The dashed box highlights the region where major R-loops start and the mutation frequency increases.

Heterogeneity of R-loop termination points. Like those of the Sγ3 and Sγ2b regions, the Sμ R-loops are heterogeneous in their termination points. Many R-loops terminate within the core Sμ repeats. We have not been able to quantify the abundance of R-loops within the core region relative to the region upstream or downstream because PCR through the core Sμ is only successful when the primers are close to the borders of the core Sμ repeats. But using native primer-converted primer pairs, we found abundant R-loops in the core Sμ region (see Table S1 in the supplemental material). The termination points of R-loops that extended downstream of the core Sμ repeats resided within the region where the G density was declining but was still higher than that of random DNA (Fig. 5). This finding was similar to those at the Sγ3 and Sγ2b repeats (9), and it is consistent with the emerging view that clusters of G's (and, indirectly, overall G density) are important for R-loop formation and stability.

Collapsed R-loops and the location of long consecutive regions of AID action. In the *ung*$^{-/-}$ *msh2*$^{-/-}$ null mice, the pattern of AID C-to-U deamination sites appears to occur in long discontinuous tracts of up to several hundred base pairs (34). This distance is far too long to be attributable to any processivity of AID along one strand, which is limited to much shorter lengths (26). The separation of DNA strands during normal transcription occurs for a length of approximately 9 bp (33), which is also far too short to account for the very long tracts of AID action. Finally, the separation of strands due to transient negative supercoiling associated with transcription would result in shorter tracts of single strandedness, and single strandedness would occur for both strands (16).

The several-hundred-base-pair tracts of AID conversion in *ung*$^{-/-}$ *msh2*$^{-/-}$ B cells on the top (nontemplate) strand (upstream, downstream, and within the core Sμ) may be attributed to R-loops, based on the findings described here. But

what could be the basis for tracts of conversion on the bottom (template) strand in these *ung*$^{-/-}$ *msh2*$^{-/-}$ B cells (34)? We believe that the exposure of single-stranded regions on the template strand of the R-loops can arise in either of two ways. The RNA of the R-loop is vulnerable to the endogenous endonucleolytic action of RNase H1 or H2 within the cell. First, partial action by RNase H may give rise to the gapped regions of single strandedness on the top strand that we have observed here for the Sμ and previously for the Sγ3 and Sγ2b regions (9) (see Fig. S4A in the supplemental material). Second, complete action by RNase H may give rise to misalignment due to the collapse of a repeat on the nontemplate strand that does not match the corresponding repeat on the template strand (see Fig. S4B in the supplemental material) (35, 37). We have previously described evidence for both of these mechanisms of RNase H-mediated exposure of regions of single strandedness on the template strand. Using purified plasmid R-loops digested in vitro, we have shown evidence for collapsed R-loops (35). The incomplete digestion of chromosomal R-loops with exogenous RNase H gives results that are quite consistent with partially digested R-loops (9, 35). Hence, both of these products of the RNase H digestion of R-loops may exist (37) (see Fig. S4A and B in the supplemental material). Either mechanism would leave long tracts of single strandedness on the template strand of the type suggested by the pattern of AID action (34). In the *ung*$^{-/-}$ *msh2*$^{-/-}$ B cells, it is noteworthy that the discontinuous tracts of AID conversion on the nontemplate strand are, on average, about twice as long as the ones on the template strand, despite the fact that the density of C's on the template strand is about 1.5-fold higher (34). The two mechanisms for template strand single strandedness suggested here would be consistent with the favoring of longer stretches on the nontemplate strand seen in the *ung*$^{-/-}$ *msh2*$^{-/-}$ B cells.

For R-loops in the repetitive zone, collapsed R-loops and the partially digested R-loops are both reasonable explanations. For R-loops that begin upstream or extend downstream of the core Sμ repeats, the DNA is not markedly repetitive. Hence, tracts of AID action are more directly explained by the partial RNase H digestion model (see Fig. S4A in the supplemental material).

Secondary structure within the G-rich DNA strand of the R-loop. The possibility has been raised that G quartets form along the G-rich DNA strand of the R-loops (4, 7). To our knowledge, there are no in vivo data for G quartets at these sites. One may have expected that G quartets would cause base pairing at other sites along the G-rich DNA strand of the R-loop and that this would cause single-nucleotide interruptions in the continuous sites of bisulfite conversion on the G-rich strand. We do not see evidence of this here or previously (35). One may wonder if the bisulfite method would be sensitive to such fine structural features along the G-rich DNA strand. Our ability to detect a small stem-loop structure on the G-rich DNA strand indicates that small regions of secondary structure within a strand are readily discernible (Fig. 3; also see Fig. S2 in the supplemental material).

ACKNOWLEDGMENTS

We thank Michael S. Neuberger for providing detailed information about the precise locations of previously described mutations (34). We also thank members of the Lieber laboratory for comments on the paper.

This work was supported by NIH grants to M.R.L. and E.S. and by grants to A.A.K. from ARC (no. 3789), LNCC (comite Tarne-et-Garonne), and Canceropole GSO (ACI 2004-2007).

REFERENCES

1. **Barreto, V. M., Q. Pan-Hammarstrom, Y. Zhao, L. Hammarstrom, Z. Misulovin, and M. C. Nussenzweig**. 2005. AID from bony fish catalyzes class switch recombination. J. Exp. Med. **202:**733–738.
2. **Bransteitter, R., P. Pham, M. D. Scharff, and M. F. Goodman**. 2003. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc. Natl. Acad. Sci. USA **100:**4102–4107.
3. **Chaudhuri, J., and F. W. Alt**. 2004. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat. Rev. Immunol. **4:**541–552.
4. **Daniels, G. A., and M. R. Lieber**. 1995. RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. Nucleic Acids Res. **23:**5006–5011.
5. **DiNoia, J., and M. S. Neuberger**. 2002. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. Nature **419:**43–48.
6. **Dunnick, W. A., G. Z. Hertz, L. Scappino, and C. Gritzmacher**. 1993. DNA sequence at immunoglobulin switch region recombination sites. Nucleic Acids Res. **21:**365–372.
7. **Duquette, M. L., P. Handa, J. A. Vincent, A. F. Taylor, and N. Maizels**. 2004. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. Genes Dev. **18:**1618–1629.
8. **Ge, X., D. S. C. Yang, K. Trabbic-Carlson, B. Kim, A. Chilkoti, and C. D. M. Filipe**. 2005. Self-cleavable stimulus response tags for protein purification without chromatography. J. Am. Chem. Soc. **127:**11228–11229.
9. **Huang, F.-T., K. Yu, C.-L. Hsieh, and M. R. Lieber**. 2006. The downstream boundary of chromosomal R-loops at murine switch regions: implications for the mechanism of class switch recombination. Proc. Natl. Acad. Sci. USA **103:**5030–5035.
10. **Huertas, P., and A. Aguilera**. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol. Cell **12:**711–721.
11. **Ichikawa, H. T., M. P. Sowden, A. T. Torelli, J. Bachl, P. Huang, G. S. Dance, S. H. Marr, J. Robert, J. E. Wedekind, H. C. Smith, and A. Bottaro**. 2006. Structural phylogenetic analysis of activation-induced deaminase function. J. Immunol. **177:**355–361.
12. **Khamlichi, A. A., F. Glaudet, Z. Oruc, V. Denis, M. LeBert, and M. Cogne**. 2004. Immunoglobulin class-switch recombination in mice devoid of any Smu tandem repeat. Blood **103:**3828–3836.
13. **Leung, H., and N. Maizels**. 1994. Regulation and targeting of recombination in extrachromosomal substrates carrying immunoglobulin switch region sequences. Mol. Cell. Biol. **14:**1450–1458.
14. **Li, X., and J. L. Manley**. 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell **122:**365–378.
15. **Lieber, M. R**. 1991. Site-specific recombination in the immune system. FASEB J. **5:**2934–2944.
16. **Liu, L. F., and J. C. Wang**. 1987. Supercoiling of the DNA template during transcription. Proc. Natl. Acad. Sci. USA **84:**7024–7027.
17. **Longerrich, S., U. Basu, F. Alt, and U. Storb**. 2006. AID in somatic hypermutation and class switch recombination. Curr. Opin. Immunol. **18:**164–174.
18. **Luby, T. M., C. E. Schrader, J. Stavnezer, and E. Selsing**. 2001. The mu switch region tandem repeats are important but not required for antibody class switch recombination. J. Exp. Med. **193:**159–168.
19. **Masukata, H., and J. Tomizawa**. 1990. A mechanism of formation of a persistent hybrid betwen elongating RNA and template DNA. Cell **62:**331–338.
20. **Min, I. M., L. R. Rothlein, C. E. Schrader, J. Stavnezer, and E. Selsing**. 2005. Shifts in targeting of class switch recombination sites in mice that lack mu switch region tandem repeats or Msh2. J. Exp. Med. **201:**1885–1890.
21. **Min, I. M., C. E. Schrader, J. Vardo, T. M. Luby, N. D'Avirro, J. Stavnezer, and E. Selsing**. 2003. The Smu tandem repeat region is critical for Ig isotype switching in the absence of Msh2. Immunity **19:**515–524.
22. **Muramatsu, M., K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, and T. Honjo**. 2000. Class switch recombination and somatic hypermutation require activation-induced cytidine deaminase (AID), a member of the RNA editing cytidine deaminase family. Cell **102:**541–544.
23. **Muramatsu, M., V. Sankaranand, S. Anant, M. Sugai, K. Kinoshita, N. Davidson, and T. Honjo**. 1999. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J. Biol. Chem. **274:**18470–18476.
24. **Petersen, S., R. Casellas, B. Reina-San-Martin, H. T. Chen, M. J. Difilippantonio, P. C. Wilson, L. Hanitsch, A. Celeste, M. Muramatsu, D. R. Pilch, C. Redon, T. Ried, W. M. Bonner, T. Honjo, M. C. Nussensweig, and A.**

Nussensweig. 2001. AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. Nature **414:**660–665.
25. **Petersen-Mahrt, S. K., R. S. Harris, and M. S. Neuberger**. 2002. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. Nature **418:**99–103.
26. **Pham, P., R. Bransteitter, J. Petruska, and M. F. Goodman**. 2003. Processive AID-catalyzed cytosine deamination on single-stranded DNA stimulates somatic hypermutation. Nature **424:**103–107.
27. **Reaban, M. E., and J. A. Griffin**. 1990. Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. Nature **348:**342–344.
28. **Ronai, D., M. D. Iglesias-Ussel, M. Fan, Z. Li, A. Martin, and M. D. Scharff**. 2007. Detection of chromatin-associated single-stranded DNA in regions targeted for somatic hypermutation. J. Exp. Med. **204:**181–190.
29. **Selsing, E**. 2006. Ig class switching: targeting the recombinational mechanism. Curr. Opin. Immunol. **18:**249–254.
30. **Shinkura, R., M. Tian, M. Smith, K. Chua, Y. Fujiwara, and F. W. Alt**. 2003. The influence of transcriptional orientation on endogenous switch region function. Nat. Immunol. **4:**435–441.
31. **Tian, M., and F. W. Alt**. 2000. Transcription induced cleavage of immunoglobulin switch regions by nucleotide excision repair nucleases *in vitro*. J. Biol. Chem. **275:**24163–24172.
32. **Wakae, K., B. G. Magor, H. Saunders, H. Nagaoka, A. Kawamura, K. Kinoshita, T. Honjo, and M. Muramatsu**. 2006. Evolution of class switch recombination function in fish activation-induced cytidine deaminase, AID. Int. Immunol. **18:**41–47.
33. **Westover, K. D., D. A. Bushnell, and R. D. Kornberg**. 2004. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. Science **303:**1014–1016.
34. **Xue, K., C. Rada, and M. S. Neuberger**. 2006. The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in *msh2*$^{-/-}$ *ung*$^{-/-}$ mice. J. Exp. Med. **203:**2085–2095.
35. **Yu, K., F. Chedin, C.-L. Hsieh, T. E. Wilson, and M. R. Lieber**. 2003. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat. Immunol. **4:**442–451.
36. **Yu, K., F. T. Huang, and M. R. Lieber**. 2004. DNA substrate length and surrounding sequence affect the activation induced deaminase activity at cytidine. J. Biol. Chem. **279:**6496–6500.
37. **Yu, K., and M. R. Lieber**. 2003. Nucleic acid structures and enzymes in the immunoglobulin class switch recombination mechanism. DNA Repair **2:**1163–1174.
38. **Yu, K., D. Roy, M. Bayramyan, I. S. Haworth, and M. R. Lieber**. 2005. Fine-structure analysis of activation-induced deaminase accessbility to class switch region R-loops. Mol. Cell. Biol. **25:**1730–1736.
39. **Yu, K., D. Roy, F. T. Huang, and M. R. Lieber**. 2006. Detection and structural analysis of R-loops. Methods Enzymol. **409:**316–329.
40. **Zarrin, A. A., F. W. Alt, J. Chaudhuri, N. Stokes, D. Kaushal, L. DuPasquier, and M. Tian**. 2004. An evolutionarily conserved target motif for immunoglobulin class-switch recombination. Nat. Immunol. **5:**1275–1281.
41. **Zarrin, A. A., M. Tian, J. Wang, T. Borjeson, and F. W. Alt**. 2005. Influence of switch region length on immunoglobulin class switch recombination. Proc. Natl. Acad. Sci. USA **102:**2466–2470.